

# Conor R. Walker

New York Genome Center, 101 6th Ave, New York, NY 10013, USA

✉ [conor@conorwalker.net](mailto:conor@conorwalker.net) • 📞 [Conor R. Walker](#) • 📧 [conorwalker](#) • [in conor-walker](#)

## Education

---

### University of Cambridge, UK

2017–2021

PhD Computational Biology

- Thesis: Statistical analysis of short template switch mutations in human genomes
- Advisers: Nick Goldman, Aylwyn Scally, Nicola De Maio

### Newcastle University, UK

2016–2017

MSc Bioinformatics · Distinction (83%)

- Thesis: Optimising nucleic acid sequences for DNA strand displacement systems
- Adviser: Harold Fellermann

### Liverpool John Moores University, UK

2013–2016

BSc (Hons) Zoology · First-class honours (80%)

- Thesis: The genetics of insecticide resistance in the blackfly *Simulium vittatum*
- Advisers: Craig Wilding, Will Swaney

## Research experience

---

### New York Genome Center, USA

2022–Present

Postdoctoral Research Associate (Joint appointment at Columbia University)

- Developed a deep learning pipeline (PyTorch) and statistical methods for assessing privacy risks associated with single-cell RNA-Seq datasets
- Implemented multiple classification models (scikit-learn) to predict genotypes from single-cell RNA-Seq, revealing privacy risks across many publicly available datasets
- Modelled noise between bulk and single-cell RNA-Seq data using generalized linear models (Python)
- Collaboratively designed a privacy-preserving bulk RNA-Seq quantification algorithm using homomorphic encryption, and processed hundreds of bulk RNA-Seq samples for benchmarking (Snakemake & AWS)

### Columbia University - Department of Biomedical Informatics, USA

2022–Present

Postdoctoral Research Fellow

### European Bioinformatics Institute (EMBL-EBI), UK

2017–2022

Predocctoral Researcher

- Designed and implemented (C++) hidden Markov models to identify small-scale rearrangements in sequencing data, providing the first statistical method for identifying <50nt DNA rearrangements
- Applied my models to thousands of whole-genome human sequencing datasets to evaluate the evolutionary, population, cancer, and *de novo* landscape of template switch mutagenesis
- Implemented convolutional neural networks (Keras & PyTorch) to accurately detect positive selection within protein-coding sequences, validated by simulating evolution across tens of millions of genes
- Collaborated on multiple projects to characterize global SARS-CoV-2 genome evolution, analyzing thousands of samples (Python/Snakemake) and co-authoring 4 peer-reviewed papers

This work was funded by EMBL and NIHR.

### Newcastle University, UK

2016–2017

Postgraduate Student Research

- I implemented multi-objective genetic algorithms for designing and optimising DNA sequences for use in DNA computing systems

### Liverpool John Moores University, UK

2016

Undergraduate Student Research

- I identified enzymes and mutations associated with insecticide resistance in a medically-important blackfly genome using phylogenetic methods

## Publications

---

phastSim: Efficient simulation of sequence evolution for pandemic-scale datasets.

De Maio N., Boulton W., Weiguny L., **Walker C. R.**, Turakhia Y., Corbett-Detig R., Goldman N.

[PLOS Computational Biology](#), 18, e1010056 (2022).

Mutation rates and selection on synonymous mutations in SARS-CoV-2.

De Maio N., **Walker C. R.**, Turakhia Y., Lanfear R., Corbett-Detig R., Goldman N.

[Genome Biology and Evolution](#) 13, evab087 (2021).

Short-range template switching in great ape genomes explored using pair hidden Markov models.

**Walker C. R.**, Scally A., De Maio N., Goldman N.

[PLOS Genetics](#) 17, e1009221 (2021).

Stability of SARS-CoV-2 phylogenies.

Turakhia Y., De Maio N., Thornlow B., Gozashti L., Lanfear R., **Walker C. R.**, Hinrichs A. S., Fernandes J. D., Borges R., Slodkowitz G., Weiguny L., Haussler D., Goldman N., Corbett-Detig R.

[PLOS Genetics](#) 16, e1009175 (2020).

A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages.

Dellicour S., Durkin K., Hong S. L., Vanmechelen B., Martí-Carreras J., Gill M., Meex C., Bontems S., André E., Gilbert M., **Walker C. R.**, De Maio N., Faria N., Hadfield J., Hayette M., Bours V., Wawina-Bokalanga T., Artesi M., Baele G., Maes P.

[Molecular Biology and Evolution](#) 38, 1608–1613 (2020).

Masking strategies for SARS-CoV-2 alignments.

De Maio N., **Walker C. R.**, Borges R., Weiguny L., Slodkowitz G., Goldman N.

[virological.org/t/masking-strategies-for-sars-cov-2-alignments/480](https://virological.org/t/masking-strategies-for-sars-cov-2-alignments/480) (2020).

Issues with SARS-CoV-2 sequencing data.

De Maio N., **Walker C. R.**, Borges R., Weiguny L., Slodkowitz G., Goldman N.

[virological.org/t/issues-with-sars-cov-2-sequencing-data/473](https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473) (2020).

## Selected presentations

---

RECOMB 2023 · Istanbul, Türkiye

Apr. 2023

Poster: Private information leakage in single-cell omics data

**Walker C. R.**, Gürsoy G.

Rocky 2022 · Colorado, US

Dec. 2022

Poster: Private information leakage in single-cell omics data

**Walker C. R.**, Gürsoy G.

ISMB/ECCB 2021 · Remote

Jul. 2021

Talk: Accurate detection of positive selection using convolutional neural networks

**Walker C. R.**, De Maio N., Goldman N.

SMBE 2021 · Remote

Jul. 2021

Talk: Accurate detection of positive selection using convolutional neural networks

**Walker C. R.**, De Maio N., Goldman N.

SMBE 2019 · Manchester, UK

Jul. 2019

Poster: Short template switches explain mutation clusters in the human genome

**Walker C. R.**, De Maio N., Goldman N.

NIHR BRC Annual Research Day · Cambridge, UK

Jul. 2018

Talk: Investigating short template switch mutations in humans and model organisms

**Walker C. R.**

## Teaching

---

### Students supervised

- |   |      |
|---|------|
| Marvin Limpijankit · Undergraduate student  | 2022 |
| • Project: Classifying somatic mutations using machine learning   |      |
| Will Lounsbery-Scaife · Undergraduate student   | 2022 |
| • Project: Deep learning models of single-cell genotype-phenotype relationships                         |      |
| Fatma Rabia Fidan · Masters student   | 2021 |
| • Project: Statistically identifying <i>de novo</i> template switch mutations in trios of human genomes |      |
| Shayesteh Arasti · Masters student  | 2021 |
| • Project: Convolutional neural networks for predicting positive selection                              |      |
| Viacheslav Vasilev · Undergraduate student  | 2020 |
| • Project: Convolutional neural networks for predicting positive selection                              |      |
| William Xu · Undergraduate student  | 2020 |
| • Project: Convolutional neural networks for predicting positive selection                              |      |

### Teaching assistant

- |  |      |
|--|------|
| Medics to Coders · University of Cambridge, UK                                     | 2018 |
| • I delivered practical sessions on programming at the School of Clinical Medicine |      |

## Honours and awards

---

- |  |           |
|--|-----------|
| ISMB/ECCB 2021 EvolCompGen - Best Talk Award                                   | Jul. 2021 |
| EMBL-EBI PhD Seminar Day - Best Talk Award                                     | May 2019  |
| MSc Bioinformatics Prize - highest overall grade on my master's degree program | Oct. 2017 |
| BSc Zoology Prize - highest overall grade on my bachelors's degree program     | Sep. 2016 |

## Other scientific activities

---

### Committees

- |   |           |
|---|-----------|
| Student Representative · EMBL-EBI, UK   | 2018–2020 |
| • Attended meetings with senior management from across all EMBL outstations to discuss issues impacting the graduate community at EMBL  |           |
| • Organised a variety of academic and social events within the institution  |           |
| Graduate Student and Post-Doc Forum member · University of Cambridge, UK  | 2018      |
| • I worked as part of a group of students and postdocs to devise strategies for creating a more productive and positive experience for graduate students in the School of Life Sciences |           |

### Memberships

International Society for Computational Biology  
 Society for Molecular Biology and Evolution  
 Society for the Study of Evolution

### Reviewer for

Nature Medicine  
 Molecular Biology and Evolution  
 BMC Bioinformatics  
 RECOMB (Research in Computational Molecular Biology)

## Skills

---

**Programming** Python, Bash, C++, R, AWK, LaTeX

**Libraries** Numpy, SciPy, Keras, TensorFlow, PyTorch, pandas, matplotlib, seaborn, Dask, Numba, pysam, scikit-allel, Scanpy, scvi-tools

**Coding practices** Git, Snakemake, Nextflow, unit testing, Docker, Singularity, Jupyter, Vim

**High performance computing** Slurm, LSF

**Cloud computing** AWS, Google Cloud

**Operating systems** Linux, MacOS